图书情报工作
第 63 卷 第 1 期 2019 年 1 月

科研数据生命周期研究和数据知识库理论架构

■ 杨乐[1] 颜石磊[1] 李洪波[2]
[1] 温州肯恩大学 温州 325060 [2] 温州医科大学 温州 325035

摘要：［目的/意义］推进国内科研数据管理体系的发展，明确三大研究内容和数据知识库的基础理论研究。［方法/过程］采用实证研究方法，分析对比国内外科研数据管理的研究现状，详细探讨科研数据管理领域的基础研究和技术论证。［结果/结论］明确指出构建科研数据管理体系的基础研究三大内容，即为科研活动周期和数据生命周期、数据知识库的理论框架和流程研究、数据知识库的技术论证。同时提出管理体系的系统建设方案和机制建设方案。

关键词：科研数据管理 数据知识库 数据生命周期
分类号：G250
DOI：10.13266/j.issn.0252－3116.2019.01.013

1 引言

近年来，科研数据管理在国外学术界和高校图书馆得到迅猛发展。2015 年美国新媒体联盟（New Media Consortium）和高校教育促进会（EDUCAUSE Learning Initiative）联合发布的《地平线报告》将科研数据管理列为高校图书馆未来发展的重要趋势之一[1-2]。2016 年美国大学和研究图书馆学会（Association of College & Research Libraries）发布图书馆十大主流研究趋势，排名第一的便是“科研数据服务”[2-3]。在协同研究、交叉学科、信息化技术快速发展的背景下，如何对大量的科研数据进行保存、发现、引用、再利用和长期维护等，已经引起学术界的高度重视[4]。

面对我国学术活动日趋活跃，各领域基础研究和应用研究在很多方面已经开始引领国际发展的大背景下，国务院办公厅于 2018 年 3 月 17 日印发《科学数据管理办法》，正式要求政府预算资金支持开展的科研活动，需要对科研数据进行合理的采集生产、加工整理、开放共享和管理使用[5]。在这种国际大环境和国务院办公厅出台政策的大背景下，国内高校科研数据管理服务却仍然停留在探讨论证的阶段，数据生命周期（Data Lifecycle）尚未建立有效的本土化模型，科研数据知识库的基础论证和研发尚未起步。如何从基础论证阶段开始，建设好科研数据管理的服务体系、建立本土化的数据生命周期模型、研究各种科研活动的流程、建立完整的科研数据管理大框架、探索开发完整的数据知识库系统等，已成为各大高校和科研机构必须面对的问题。这不仅仅关系到高校和科研机构自身的战略规划，同时也关系到国内科学研究的有序发展。

2 研究现状

2.1 国外研究现状

美国高校和科研机构的数据管理服务在经费投入、从业资质要求、岗位设置、技术要求等方面都设置了较高的标准。从学术活动的初始阶段课题申请开始，高校图书馆数据管理团队就提供了嵌入式的咨询服务[6]。利用科研数据管理计划平台（如加州数字图书馆开发的 DMP Tool[7]），并结合一系列图情专业人员的专长，如数据格式、数据存档、数据监护、元数据配置方案等，指导课题申请者和研究者撰写符合各大基金会要求的数据管理方案。与咨询服务配套的是科研数据管理的培训模式，多数高校图书馆科研数据管理员都会以专题讨论、讲座、在线学习等方式，对科研人员，教授，以及本校研究生进行数据素养（Data Literacy）和科研数据生命周期（Data Lifecycle）的培训[7-8]。

密歇根大学（University of Michigan）图书馆情报学学者 K. Akers 等[9]着重梳理了美国本土八大高校图书馆传统强校的科研数据服务、管理、评估、传统机构知识库构建、数据知识库构建的时间表，如图 1 所示：

作者简介：杨乐（ORCID：0000－0001－6504－9487），副研究馆员，硕士；颜石磊（ORCID：0000－0003－1263－7546），硕士；李洪波（ORCID：0000－0003－0988－0385），硕士，通讯作者，E-mail：lihb@wmu.edu.cn。

收稿日期：2018－10－17 修回日期：2018－11－23 本文起止页码：91－97 本文责任编辑：徐健

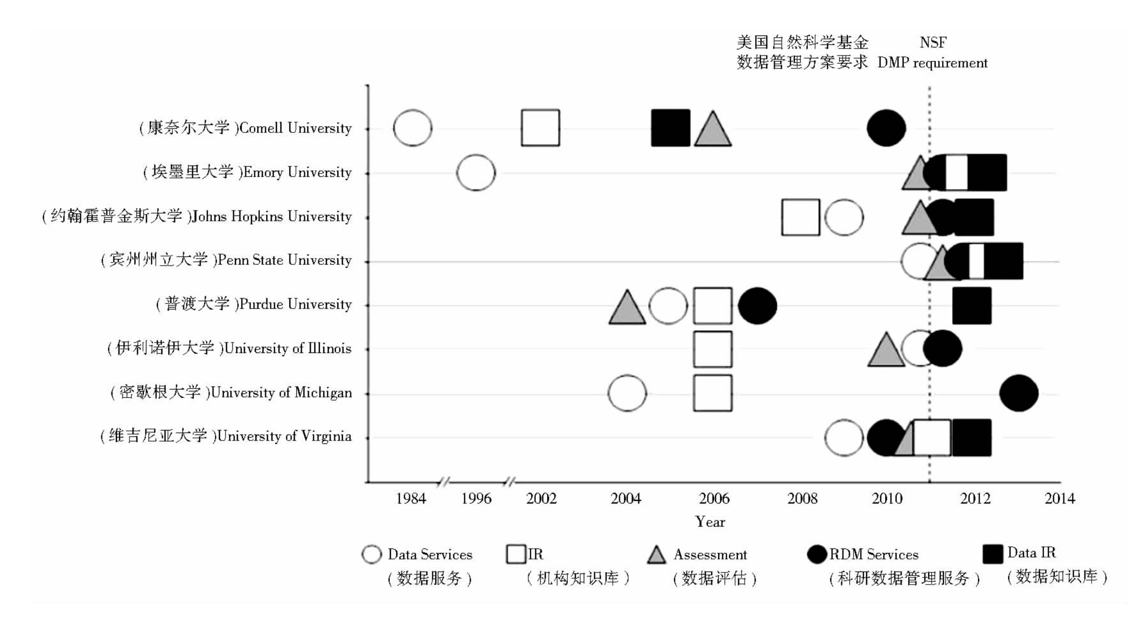


图1 美国八大高校图书馆科研数据服务的时间表[9]

从图1可以看出，康奈尔大学图书馆早在1984年就开始进行科研数据管理的服务工作，而埃默里大学则在1996年开始。国立卫生研究院（NIH）于2003年2月26日开始要求50万美元以上的科研项目必须有配套的科研数据管理计划书和落实方案。在那之后，科研数据管理和数据知识库开始迅速发展。2011年1月18日，国家科学基金会（NSF）开始要求所有的科研基金申请书必须包含有配套的科研数据管理办法和执行方案。在此之后，各大科研能力较强的高校图书馆开始了数据管理的普及工作。

2013年2月22日，奥巴马政府签署了一个科技政策备忘录，要求所有获得联邦基金资助的科研任务在结束之后的规定时间内，必须在开放存取（Open Access）的机构知识库或者数据知识库上无条件公开出版原始科研数据。在这一政策落实之后，所有的联邦科研基金会都陆续颁布了对科研数据管理的具体细则。紧接着，由美国各大院校和科研机构图书馆牵头，全方位启动了科研数据管理方面的相关工作。具备科研能力的学术型图书馆率先推出了科研数据管理的服务和咨询，并且在现有的机构知识库平台上进行初步的科研数据的存储和发表，同时通过馆际合作开展科研数据管理系统的开发工作。

在数据知识库系统方面，首先，最具特色的开源系统有哈佛大学的Dataverse和北卡罗来纳大学系统的iRODS；其次，具有区域独特性和个性化特色的数据知识库如普渡大学的PURR、加州大学系统联合开发的DataONE（现版本为Dash）、密歇根大学开发的专门针对社会人文科学的ICPSR[10]等；第三，也有众多高校选择用传统的机构知识库DSpace、Fedora等系统来存储和发表科研数据[11]；再者，现有期刊文献中尚未有所涉及的最新出现的下一代机构知识库系统Samvera，涵盖了传统功能和科研数据功能，由Fedora为后台，加上专门的前台应用Hydra，并以此为基础开发出一系列的各种独立的多功能平台Avalon、Hyku、Hyrax等以支持不同数据文件格式。

2.2 国内研究现状

2018年3月17日，国务院办公厅印发《科学数据管理办法》，正式要求政府预算资金支持开展的科研活动，需要对科研数据进行合理的采集生产、加工整理、开放共享和管理使用[5]。现阶段，国内在各层级科研基金资助方面还没有数据管理方案和数据管理执行层面上的要求。多数国内高校图书馆和科研机构都较少涉及到科研数据管理或者数据管理咨询服务的领域。目前，国内的期刊文献探讨了国外高校的数据管理理念、模式和系统的比较[12]，并且着重讨论了对国内高校发展的启示[13-15]，但在推广和执行科研数据管理实践方面进展不大。各大高校的管理层和学术团队对科研数据管理方面的工作并没有很强的支持力度，在经费投入、从业资质、岗位设置、技术要求等各方面都要求不高[2,12]。

对科研数据生命周期的研究，是科研数据管理的前期重点。只有在本土化的基础上建立起适用于国内科研大环境的数据生命周期模型，或者是各个高校针对本校的学科特长建设起来的数据生命周期模型，对

科研数据管理的服务工作和系统建设才有针对性作用[3]。数据知识库的结构流程建立和功能设置,亦是根据科研数据生命周期来进行。国内对科研数据生命周期的研究并不多见,仅有的期刊文献也仅是停留在对国外相关成果的比较和讨论,同时针对性地提出对国内图情方面相关建设的启示[16－17],并没有在学术成果方面体现出一套成熟的、适用于国内科研环境的数据生命周期和架构模型。

在数据知识库系统方面,相较于美国已经全面普及的传统机构知识库和数据知识库结合使用的现状[18－20],国内的高校对机构知识库和数据知识库的总体发展水平并不平衡[21]。很多大学还没有建立起传统机构知识库或者数据知识库,只有屈指可数的大学已经建立起了初步的数据知识库。比如,北京大学的开放研究数据平台、复旦大学的社会科学数据平台、香港科技大学的 DataSpace,都使用了美国的 Dataverse 开源软件。在知识库开发技术上看,国内除了现有的传统机构知识库,诸如中国科学院国家科学图书馆兰州分馆的 CSpace 和中国知网的 CIRP 等之外,并没有本土化的数据知识库系统。

3 问题陈述

美国学界的知识库系统功能和技术各有所侧重,但都有其不足的一面。Dataverse 是一个同时具有前台和后台的完整 Java 应用,但是其现有的前端并未实现真正的同后端逻辑分离,Rest API 设计也不是最佳方案,可客户化程度较低,提供一定的数据可视化服务。它的优点是在元数据配置方面的高度个性化操作,另外可以对科研数据的不同版本进行不同阶段的存档备份,流程比较完备,尤其适用于地区性的数据,能较好地整合其他现有的传统知识库软件。iRODS 采用 C/C++语言,主要针对超大型的分布式数据,处理大数据反应快速,常用于气象学数据建模、生物核酸序列分析等。缺点是用户界面不友好,因其前端应用由某些 Microservice 插件提供,使得相应的客户化和深度配置比较困难。

而在众多的传统机构知识库软件里面,Fedora 比较适合用于数据知识库的工具,在处理分布式大数据功能接近 iRODS 的前提下,其最突出的优点是有着先进完善的 Rest API 网络服务,还有包括高精度版本控制、高速缓存、支持多种数据存储技术、可插用多种用户安全识别系统、高度可扩展性架构等。在用户界面和科研数据对接的技术支持也远远超前于以上提及的 iRODS,支持更好的可视化功能,所以针对 Fedora 的前端应用开发非常热门,能适用于更广泛的学科领域。另外,一个被广泛使用的是 DSpace,其前端框架 Cocoon 早已经过时。但是正在开发的 DSpace 7 将采用全新的 Angular 2/4 来创建单页应用程序,配有全新设计的 Rest API 方案。DSpace 的缺点在于对有些数据格式版本不支持,也不支持复合数据类型,可视化手段也比较缺乏。DSpace 因而更适用于出版的文献资料。总之,不管以上的哪一种方式,都是独立于数据管理计划系统之外,没有集各功能为一体的聚合式系统。

综上所述,国外的数据知识库虽然有较完善的体系,然而并没有很好地解决整体功能上的聚合,只是片面地整合了某些系统功能,存在多个功能系统并存但无法兼容的特点[22－23]。另外,据美国学者 R. Uzwyshyn 在 2016 年做的调查发现,虽然美国已经有 74% 研究型大学提供了科研数据的系统平台,却只有 13% 使用专门的数据知识库平台,其他的高校都是使用传统机构知识库或者网站代替[24]。而国内高校的科研数据管理和数据知识库尚在基础论证阶段,最为基础的数据生命周期模型还没建立、仅有的为数不多的数据知识库采用了美国的现有技术;然而其现有技术存在技术短板,语言模块特别是元数据对国外学术检索系统的发现功能支撑不足。这种现象不止局限于国内,在一份针对中日韩三国数据知识库的调查研究中,韩国学者 S. Kim 和 W. Lee 发现虽然中日韩三国知识界的数据知识库已经占据了整个亚洲学术界的 42.2%,然而绝大多数都是传统意义上的数据库框架和机构知识库(如 DSpace)框架[25]。

国内图情研究在现阶段需要对国际上现有的数据知识库的架构和模型进行更深层次的研究,并结合国内高校研究人员对科研数据管理的本土化需求,建立本土化数据生命周期模型,在此基础上构建一套更加完善、本土化(例如:中文元数据集本土化)和国际化(例如:配置 Schema.org 或者 OAI-PMH 元数据协议以便搜索引擎检索)兼具的数据管理模型和数据知识库系统。更加完善特指利用互联网技术融合科研周期和数据生命周期的各个阶段,包括初期的科研数据管理方案系统、中期的数据仓库系统、以及后期的数据整合、出版、引用、开放获取系统,同时能够对接美国高校的科研数据发现系统或者谷歌(Google)于 2018 年 9 月 5 日发布的数据集搜索引擎(Google Dataset Search)。

4 研究内容

面对这一领域的研究现状,国内图书馆情报学界如果需要在科研数据领域取得突破,需从三个主要方面(三个阶段)开始着手:①细化的科研活动周期和科研数据生命周期模型;②数据知识库的理论框架和流程;③数据知识库的技术方案。

4.1 科研活动周期和数据生命周期模型

科研数据生命周期研究在美国学界已经由来已久,多数美国研究图书馆协会(Association of Research Libraries,ARL)成员的学术型图书馆都根据本校的学科设置而建立起一个通用的生命周期模型,用于指导校内科研数据管理服务工作的开展,同时用于指导数据知识库的开发。如图 2 所示:

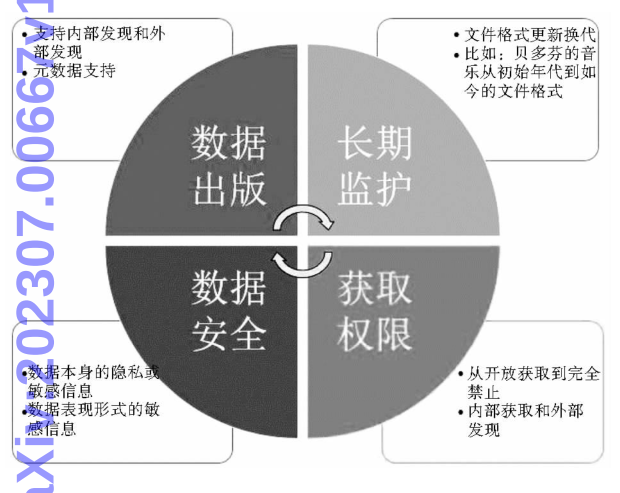


图 2 科研数据生命周期通用模型

在图 2 中可以看出,一个具有通用性的数据生命周期模型包括 4 个模块:数据长期监护保证了对数据文件完整性的长期保护和对文件格式的更新换代;数据的获取权限,指科研人员根据科研活动周期和流程,相应的选择数据开放的权限,方便校内外课题组成员,或者跨课题组合作人员的科研数据共享;数据安全指数据图书馆员根据不同级别的基金委员会或者各层级机构的相关要求,对数据集涉及到的人群身份、敏感信息等信息进行隐藏处理;数据出版则不仅涉及到各层级基金委员会对科研数据的开放获取的要求,同时要求配套的保证数据完整性的监护元数据策略和支持数据外部共享的发现元数据策略。

对细化学科的科研活动流程机制的研究以及对科研数据生命周期进行本土化的研究均为非常必要的环节,将直接影响数据管理系统的框架构建和数据知识库的理论架构。通过具体学科领域的科研活动的调查研究,研究人员可以建立起一套或多套科研活动流程和模型,基于有共通性科研活动流程的学科和有特定流程的学科,进一步深入了解科研数据在科研活动流程的生命周期,从而构建有适用性的本土化数据生命周期模型。对这两项内容的研究,需要通过调查问卷和访谈的形式,从科研人员获取第一手的反馈,同时获取各学科的实验数据格式信息。

4.2 数据知识库的理论框架和流程

第二阶段的研究内容,是能够充分研究国际上现有的数据知识库和相关配套系统处理科研数据的总体流程,解构数据知识库的总体框架。进而结合本土化的科研周期和科研数据生命周期模型,搭建数据知识库系统的结构框架和模型论证。

根据现有国际上的文献和实证案例,主要的科研数据管理体系有三大功能模块和流程:数据管理计划、科研数据存储和出版、科研数据的长期监护和整合利用(见图 3)。此体系能够覆盖科研数据的整体生命周期,然而目前国际上并没有一个数据知识库系统能够把三大功能模块进行有效的聚合,大部分都是自成一个独立的系统体系。比如,加州数字图书馆的 DMP Tool 用于数据管理计划书、Dash 用于数据存储和出版引用、Chronopolis 用于科研数据的长期维护;也有知识库能局部地融合数据存储、引用、出版和长期监护再利用,比如普渡大学的 PURR 系统,唯独缺少了数据管理计划部分。

从图 3 看出,数据管理计划平台在科研周期的开端允许数据管理人员嵌入到科研周期的前期数据管理计划中来。科研人员在平台系统注册之后可以根据不同学科的需求调取相应的数据管理计划模板,在数据管理人员的协助下撰写科研数据管理计划书并提交审核,确保科研人员在科研活动的初始阶段即有科研数据管理的提前构想和布局。而相对应学科的数据管理计划模板,则需要建立在本土化和各学科的科研数据生命周期模型的研究基础上。

在数据仓库平台和保存整合平台,数据知识库的功能设置必须满足相对应的数据生命周期的功能,允许科研人员和数据管理人员在数据知识库平台上协助工作,完成数据的提交、元数据配置、优化系统发现、设置分享和使用权限、跨学科的数据整合再利用、到最后的数据出版和引用。数据知识库平台这一系列的功能建设,都必须建立在前期的科研数据生命周期的模型研究之上。如果没有较为精准的数据生命周期模型作为理论支持,数据知识库的流程和功能建设都将无章可循。

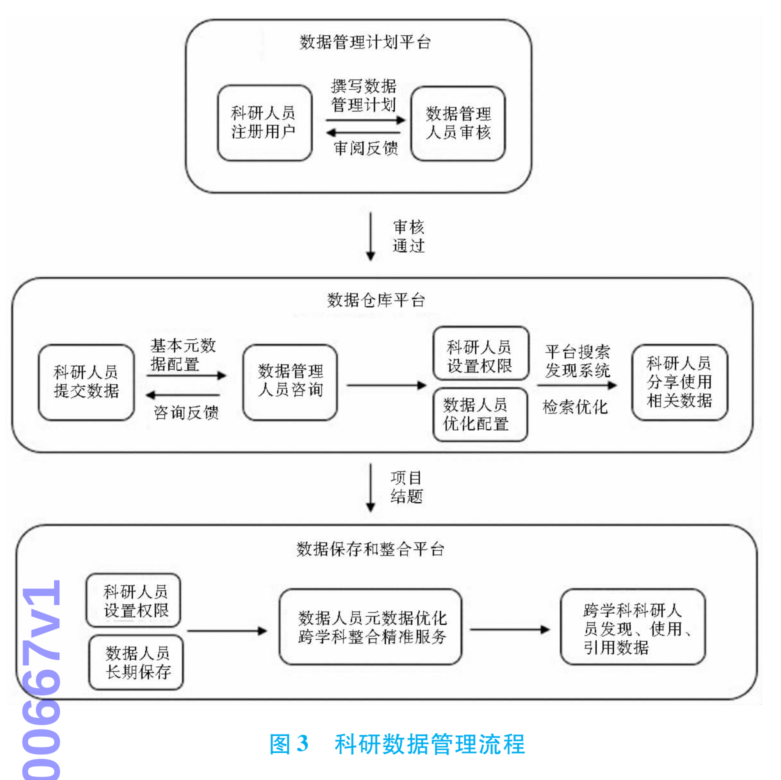


图 3 科研数据管理流程

4.3 数据知识库的技术论证

在技术层面实现一个完善的科研数据管理流程，首先要遵循 3 个原则：①充分认识和利用已有的国内外的资源，选择最适合自己的软件平台；②在软件设计方面注意功能的解耦合，使得各个模块既积极合作，又互不影响。任何一个模块的客户化和更新换代都尽量不会影响其他模块的正常工作；③软件产品技术要跟上时代的步伐，而且在技术应用层面要争取有所创新。

在现有的软件平台基础上，深入研究现有数据知识库的原理技术，如善于处理分布式大数据功能的 Fedora 和 iRODS 等开源技术、前后端逻辑达成一致的 Dataverse、或者采用 Angular 2/4 来创建单页应用程序并配有全新设计 Rest API 的 DSpace 7 版本。在技术上探索三大系统功能模块的最合理数据存储架构和科研数据管理业务逻辑在技术上的实现和优化。最后在通过充足的实证数据、理论和技术基础上，构建一个完善的科研数据管理系统模型。

在最重要的知识库后端平台方面，需要在设计上融合先进的理念，诸如以关联数据(linked data)和语义网络(semantic Web)为导向的后端平台。同时后端平台也必须支持高度客户化的元数据格式和具有处理任何格式文件的能力，从而能针对不同学科领域制定相应的科研数据管理系统，在本土化的研究方面也可提供更多的灵活性。客户化后的后端平台将接合主流用户认证授权系统(如 LDAP 系统)，并提供多种存储选择模式，建立以元数据为基础的，快速的中英文的全文搜索服务，并连接世界主流科研数据库和数据集发现系统(如 Google Dataset Search)以方便科研数据的发现、发表和利用。

最后需要解决的技术方案则是用户平台的合理化和前后台元数据的优化策略。用户平台可以分为三个不同组别的用户：外在的科研数据使用者、科研数据管理人员和科研人员。用户平台将给科研人员、信息数据使用人员以及普通用户提供快捷方便的单网页应用和移动应用，以单网页应用和移动应用为特色的数据知识库管理平台。这个前台界面只能通过设计良好的 Rest API 网络服务交流，真正做到前台后端完全分离解耦。

5 建设方案

通过第一阶段对科研数据和各学科的科研人员的学术行为进行问卷调查和访谈，深入了解和掌握科研数据类型、格式、应用范围、使用方法、数据生命周期、以及科研人员的研究方法。以此为基础，图情学界和业界的研究人员方可开始对数据知识库的建设方案进行细致规划。为了有效促进系统建设的进行，知识库的建设方案从整体上可以分为两步：系统建设方案和机制建设方案。

5.1 系统建设方案

以图 3 为例，流程图的第一部分为数据管理计划，应该满足不同学科的科研人员根据学科需求，调取相应的计划书模板进行数据管理计划的制定。数据管理人员对计划书进行审核和反馈，帮助科研人员完善数据管理计划书，才能在科研活动中更合理地管理科研数据。在流程图的第二部分，图书馆数据管理人员需要在数据库层面建设一套标准化的元数据方案覆盖各个学科，同时兼具元数据元件的可扩展性以满足具体学科的要求。科研人员提交科研数据并进行元数据的初始配置，数据管理人员再进行后期的标准化作业，确保科研数据能够被搜索引擎有效检索，方便科研数据被循环利用。在流程图的第三部分，数据管理人员对科研数据进行长期监护作业以确保科研数据的完整

性。同时进行科研数据跨学科整合和确保互操作性，对跨学科的科研人员提供精准的个性化服务，帮助科研人员实现对现存科研数据的跨学科检索、使用和引用。

5.2 机制建设方案

一个完善的科研数据管理范式(包括数据知识库在内)，必须有相应的完整机制进行配套，从而更好地服务于校内或者校际的学术科研工作。具体的机制需要遵从几个原则：①所有提交进入数据知识库的科研数据，都必须有配套的通过审核的数据管理计划。数据管理人员和科研人员协同贯彻落实数据管理计划。②在科研项目进行中和结束之后，科研人员可以在系统上进行共享权限的更改，以确保科研项目的有序进行。在不违反基金委员会规章制度或者不违反相关机构规定的前提下，科研人员可以随时要求删除科研数据。③在科研项目进行中和科研项目结束之后，数据管理人员以优化检索和发现机制为原则，对元数据进行标准化配置和修改。④根据单位的实际情况，科研人员申请科研基金的同时，应把相应的数据管理费用放进预算申请当中。科研数据管理系统亦因此而得到持续稳定的资金支持、维护、和升级。

6 结语

笔者通过介绍国内外学界和业界在科研数据管理领域的研究现状，分别陈述了国内外科研数据管理体系的发展问题，包括科研数据生命周期的研究和数据知识库的构建等方面的发展不足之处。笔者明确指出为了推进国内科研数据管理领域的发展所需要面临的三大研究内容，即为科研活动周期数据生命周期、数据知识库的理论框架和流程研究、数据知识库的技术论证。通过细化的深入研究，掌握科研数据管理体系的基础，进而系统的开展数据知识库的开发和科研数据管理的体系构建，包括系统建设和机制建设。科研数据管理体系和数据知识库体系的构建和优化、包括科研数据管理的服务体系的推广和完善，都是一个长期的且需要多个领域共同推进的过程。本文所提供的科研数据管理系统和体系构建，以科研流程和数据生命周期为切入点，脚踏实地地解决国内科研数据管理最根本的问题，建立一套符合国内科研环境和科研流程的数据周期模型，不仅为科研数据管理体系奠定基础，也是为高校图书馆更好地给科研团队推广数据管理服务奠定基础。

本文同时结合国外现有的技术特点，探索最适合开发面向国内的计算机语言环境、框架结构、本土化流程、元数据配置、制度建设等一系列从无到有的体系。从根本上解决了国外的科研数据知识管理系统的聚合问题，兼容了数据管理计划阶段、数据管理服务阶段、数据管理存储阶段、数据管理的出版阶段和数据管理的长期监护阶段的管理体系。更好地响应国务院办公厅对科研数据管理的号召，帮助科研人员更加有序合理的计划、管理、循环利用和出版科研数据。

参考文献：

[1] NEW MEDIA CONSORTIUM. EDUCAUSE. Horizon report [EB/OL]. [2018-05-21]. http://cdn.nmc.org/media/2015-nmc-horizon-report-library-EN.pdf.

[2] 国华. 高校图书馆科研数据管理服务研究[J]. 图书馆学刊, 2017(9): 81-84. DOI: 10.14037/j.cnki.tsgxk.2017.09.019.

[3] 陈媛媛,柯平. 大学图书馆科研数据服务模型研究[J/OL]. 情报理论与实践. [2018-05-21]. http://kns.cnki.net/kcms/detail/11.1762.G3.20171204.1444.008.html.

[4] Association of College & Research Libraries. 2016 ACRL academic library trends and statistics for Carnegie Classifications: associates of arts colleges baccalaureate colleges master's college and institutions doctorate granting institutions [M]. Chicago: ACRL, 2016.

[5] 国务院办公厅. 科学数据管理办法[EB/OL]. [2018-05-21]. http://www.gov.cn/zhengce/content/2018-04/02/content_5279272.htm.

[6] 张莎莎,黄国彬,邸弘阳. 美国高校图书馆科研数据管理服务研究[J]. 图书馆杂志, 2016 (7):59-66. DOI: 10.13663/j.cnki.lj.2016.07.008.

[7] 吴新年. 学术图书馆的科研数据管理服务[J]. 情报资料工作, 2014(5):74-78.

[8] 邢文明,吴方枝,司莉. 高校图书馆开展科研数据管理与共享服务调查分析[J]. 图书馆论坛, 2013(6):19-25.

[9] AKERS K, SFERDEAN F C, NICHOLLS N, et al. Building support for research data management: biographies of eight research universities [J]. International journal of digital curation, 2014, 9 (2):171-191.

[10] 王明明,王卷乐,赵强,等. ICPSR 科学数据中心的建设经验与启示[J]. 中国科技资源导刊, 2017 (6):100-107.

[11] AMORIM R, CASTRO J, DA SILVA J, et al. A comparison of research data management platforms: architecture, flexible metadata and interoperability [J]. Universal access in the information society, 2017, 16 (4):851-862.

[12] 陈和. 刍议国内高校科学数据知识库的构建[J]. 数字图书馆论坛, 2017(12):45-51.

[13] 陈媛媛,柯平. 高校图书馆科研数据服务研究综述[J]. 图书馆工作与研究, 2017 (12):67-72.

[14] 邵媛,滕文静. 面向科研数据管理的高校图书馆聚合式学科服务体系构建[J]. 图书馆学刊, 2017 (2):74-77.

[15] 王璞. 可持续发展的科研数据管理信息基础设施研究[J]. 图

书馆建设,2016(8):44 –48.
[16] 聂峰英. 基于科研团队生命周期的科研服务模型设计[J]. 图书馆, 2016(9):87 –91.
[17] 李航. 基于数据生命周期模型的学术图书馆科研数据管理体系构建[J]. 图书馆学刊, 2016,(12):19 –33.
[18] 杨志伟,卫军朝. 基于 Data Curation 的机构库建设研究——以约翰霍普金斯大学 Data Conservancy 项目为例[J]. 图书馆学研究, 2016 (7):55 –61.
[19] TRIPATHIS M, SHUKLA A, SONKER S. Research data management practices in university libraries: a study [J]. Journal of library & information technology, 2017, 37 (6):417 –424.
[20] ROOS K, MIAS E, VAN ROOYEN J. Suggesting an institutional data repository for the University of Cape Town [R/OL]. [2018 –05 –21]. DOI: 10.5281/zenodo.263823.
[21] 刘霞,饶艳. 高校图书馆科学数据管理与服务初探——武汉大学图书馆案例分析[J]. 图书情报工作,2013,57(6):33 –38.
[22] PATEL D. Research data management: a conceptual framework [J]. Library review, 2016, 65 (4/5):226 –241.
[23] MAKANI J. Knowledge management, research data management, and university scholarship: towards an integrated institutional research data management support-system framework [J]. VINE, 2015, 45 (3):344 –359.
[24] UZWYSHYN R. Research data repositories: the what, when, why, and how [EB/OL]. [2018 –05 –21]. https://digital.library.txstate.edu/handle/10877/7597.
[25] KIM S, LEE W. Global data repository status and analysis: based on Korea, China and Japan [J]. Library hi tech, 2014, 32 (4): 706 –722.

作者贡献说明:
杨乐:论文研究框架指导,论文思路修正;
颜石磊:网络前期调研,文献查阅总结;
李洪波:选题确定,论文撰写修改。

Theoretical Study of Research Data Lifecycle and Data Repository

Yang Le[1] Yan Shilei[1] Li Hongbo[2]
[1] University Library, Wenzhou-Kean University, Wenzhou 325060
[2] University Library, Wenzhou Medical University, Wenzhou 325035

Abstract: [**Purpose/significance**] The paper aims to promote the systematic development of research data management, with an emphasis on the three main aspects of fundamental research and theoretical study. [**Method/process**] The research utilizes empirical study to investigate and discuss the status quo of research data management from both the domestic and international perspectives, compares and analyzes in detail regarding the fundamental research and technological framework. [**Result/conclusion**] The result concludes three aspects of fundamental research for constructing a system of research data management, including research activities and data lifecycles, systematic workflow and theoretical framework, and technological demonstration. The conclusion also proposes strategies for system and policy development.

Keywords: research data management data repository research data lifecycle